

Final Technical Report

Award Number: 1434-HQ-97-GR-03198

Applicant: The Regents of the University of California
Sponsored Projects Office
Room 336 Sproul Hall
Berkeley, CA 94720

Principal Investigator(s): Dr. Barbara Romanowicz
Berkeley Seismological Laboratory
University of California at Berkeley
Berkeley, CA 94720-4760
Telephone Number: (510) 643-5690
FAX Number: (510) 643-5811
Email Address: barbara@seismo.berkeley.edu

Project Title: Operation of the Northern California Earthquake Data Center

Element Designation: V

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number 1434-HQ-97-GR-03198. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Abstract

We have continued the operation of the Northern California Earthquake Data Center, which serves as an on-line archive for earthquake related data collected in northern and central California, such as data from the Berkeley Digital Seismic Network, the Northern California Seismic Network and the Bay Area Regional Deformation array. The data are available to outside users through Internet. In 1997-98, we have purchased and installed the first elements of a new mass-storage device, to allow the expansion of the data base beyond the 600Gbytes capacity initially available and exceeded at the end of 1997.

Non-Technical Summary

We have continued the operation of the Northern California Earthquake Data Center, which serves as an on-line archive for earthquake related data collected in northern and central California, such as data from the Berkeley Digital Seismic Network, the Northern California Seismic Network and the Bay Area Regional Deformation array. The data are available to outside users through Internet. In 1997-98, we have purchased and installed the first elements of a new mass-storage device, to allow the expansion of the data base beyond the 600Gbytes capacity initially available and exceeded at the end of 1997.

Introduction

The Northern California Earthquake Data Center, a joint project of the Berkeley Seismological Laboratory and the US Geological Survey at Menlo Park, serves as an "on-line" archive for various types of digital data relating to earthquakes in central and northern California. The NCEDC is located at the Seismological Laboratory, and has been accessible to users via the Internet since mid-1992.

The primary goal of the NCEDC is to provide a stable and permanent archival and distribution center of digital waveforms and parametric information for earthquakes in northern and central California. The principal networks contributing seismic data to the data center are the Berkeley Digital Seismic Network (BDSN) operated by the Seismological Laboratory, and the Northern California Seismic Network (NCSN) short-period network operated by the USGS. The collection of NCSN digital waveforms dates from 1984 to the present, and the BDSN digital data dates from 1987 to the present.

Datasets

The initial phase of archiving the NCSN earthquake seismograms from 1984 through the present and BDSN data from 1987 through the present is basically complete. All NCSN data tapes have been read and data for all local and regional events has been loaded onto the NCEDC. A list of teleseismic events recorded by the NCSN is now available on the NCEDC, and the NCSN is determining which of those event have sufficient data to archive at the NCEDC. The 16-bit BDSN from 1987-1991 has been converted to MiniSEED data and is now on-line.

BDSN Seismic Data

The archival of current BDSN seismic data is an ongoing task. BDSN data is telemetered from 26 seismic dataloggers in real-time to the BSL, where it is written to disk files. Each day, a daily extraction process creates a daily archive by retrieving all continuous and event-triggered data for the previous day. The daily archive is run through quality control procedures to correct any timing errors, triggered data is reselected based on the REDI, NCSN, and UCB earthquake catalogs, and then archived at the NCEDC.

NCSN Seismic Data

NCSN waveform data continue to be collected and processed by the USGS at Menlo Park, and shipped to the datacenter on exabyte tapes. The NCSN has developed procedures to send event waveform files from Menlo Park to the NCEDC via the Internet, but was delayed by TCP/IP software problems on the Vax computers at Menlo Park. New software has been acquired that fixes the problem, and we should be able to initiate that process shortly. Parametric information, such as event catalogs and phase readings from both the BDSN and NCSN networks are automatically updated on the NCEDC on a daily basis. Figure 2 shows the number of events and event waveform files archived at the NCEDC for each year.

New problems were encountered this year with archiving NCSN seismograms at the NCEDC. In 1995, the USGS started to create seismograms files that included data from different sources and with differing data formats (16 bit data and 24 bit data). In spring 1996, the USGS addressed this problem by developing a waveform header that would be included on all seismogram files that contained data from sources other than the traditional 16-bit CUSP digitizers. Since data tapes had already been written which did not include the waveform header, the data import procedures at the NCEDC were modified to reject any seismogram

file that contained data from multiple sources and did not have the required waveform headers. The NCEDC has since been using these modified data import procedures. The USGS re-wrote the seismograms with waveform headers, and these events were loaded at the NCEDC. However, it recently came to our attention that the USGS was still failing to create waveform headers on some waveform files that contain data from multiple sources. These events have not been loaded at the NCEDC. The USGS and the NCEDC are working to resolve this problem.

BDSN Electro-Magnetic Data

The NCEDC continues to archive and process electric and magnetic field data acquired from 3 dataloggers at two sites (SAO and PKD1). The 3 channels of magnetic and 2 channels of electric data at each site are telemetered in real-time along with seismic data to the Seismological Laboratory, where they are processed and archived at the NCEDC in a similar fashion to the seismic data. The 40 Hz data channels will remain on-line for ~ 6-12 months, while the 1 Hz data will remain on-line permanently. Using programs written by Dr. Martin Fullerkrug at the Stanford University STAR Laboratory, the NCEDC is computing and archiving magnetic activity and Schumann resonance analysis with this dataset.

A new datalogger was added at PKD1 to acquire 8 channel of low frequency long baseline electric field data from an ongoing project by Dr. Steve Park of UC Riverside. This data is acquired and archived in an identical manner to the other electric field data at the NCEDC.

GPS Data

The NCEDC continues to expand its archive of gps data through the BARD (Bay Area Regional Deformation) network of continuously monitored GPS receivers in northern California. The NCEDC GPS archive now includes 40 sites in northern California. There are ~ 20 core BARD sites owned and operated by UC Berkeley, LLNL, USGS, UC Davis, Trimble Navigation, and Stanford. Data from the other northern California sites is collected from sites operated by JPL, the US Coast Guard, and Scripps Institute of Oceanography.

Most of the Berkeley BARD sites are co-located with seismic stations, and data from these sites are acquired in real-time using shared frame relay telemetry link. The remaining Berkeley stations use dedicated frame relay and/or spread spectrum radio to provide data in real-time to UC Berkeley, and are automatically processed and archived at the NCEDC on a daily basis. Data from the USGS sites are downloaded by the USGS and transferred to the NCEDC on a daily basis, and is automatically archived by the NCEDC. The other sites are automatically acquired from their respective operators, and are archived by the NCEDC.

Unocal Seismic Data

The Unocal Corporation operates a micro-seismic monitoring network in the Geysers regions of northern California, and has agreed to release six years of triggered waveform data (1987-1994) for archival and distribution at the NCEDC. This dataset represents roughly 100,000 events that were recorded by the Unocal Geysers network, and is currently available via research accounts at the NCEDC. Although no parametric information such as hypocenters or phase readings is currently being made available by Unocal, several scientists have already deemed this dataset as a useful addition to the NCEDC. The data are accessible via research accounts at the data center.

Parkfield High Resolution Seismic Network

Event seismograms from the Parkfield High Resolution Seismic Network from 1997 to 1992 have been loaded onto the NCEDC in their raw SEG-Y format. A number of events have faulty timing due to the lack or failure of a precision time source for the network. Due to funding difficulties, there is currently no ongoing work to correct the timing problems, create a catalog for these events, or to provide the NCEDC with additional data to archive. If these funding issues are resolved, we plan to resolve the timing problems in the data, create a catalog of events, archive subsequent data, and re-archive the data in MiniSEED format.

CNSS Worldwide Earthquake Catalog

The NCEDC, in conjunction with the Council of the National Seismic System (CNSS), is producing and distributing a world-wide composite catalog of earthquakes based on the catalogs of the national and various US regional networks. Each network updates their earthquake catalog on a daily basis at the NCEDC, and the NCEDC constructs a composite world-wide earthquake catalog by combining the data, removing duplicate entries that may occur from multiple networks recording an event, and giving priority to the data from each network's "authoritative region". The catalog, which includes data from 10 regional and national networks, and is available for use at the NCEDC, and is made available to anyone over the Internet.

New Hardware

The NCEDC archive exceeded the capacity of its two 300 GByte Sony WORM jukeboxes this year. All 100 slots in the Sony jukeboxes are filled with data platters, and during the year ~ 15 data platters containing older EM and raw GPS data were removed from the jukebox in order to make room for additional newer data. Partial funding for a new mass storage system was made available from the USGS in mid-year, but the purchase of the new mass storage system was deferred in an attempt to acquire higher density drives, which ultimately were not available in time.

The new mass system was purchased in late spring 1998, and is comprised of a Sun Ultra 450 computer with 108 GBytes of raw disk, a 1.3 Tbyte DISC 517 slot jukebox with two 2.6 GByte Magneto Optical (MO) drives, two 11-slot AIT tape jukeboxes which holds 25 GBytes per tape, and the SAM-FS hierarchical file system management software. Only two MO drives and minimal media were purchased since it is assumed that the jukebox will be upgraded to a 4 drive (2.6 TByte capacity) next year with the higher capacity 5.2 Gbyte MO drives. The mass storage system can additionally be upgraded to 100 slots (5.2 TByte) capacity with the addition of a second media picker, drives, and media cells. One of the AIT tape jukeboxes is designated for standard filesystem backups, and the second one is configured for use the SAM-FS software for secondary storage of archived data.

The SAM-FS software provides automated and transparent data archival and retrieval of files stored on removable media, and is used to administer and control the DISC MO and AIT tape jukeboxes. The SAM-FS software can be configured to automatically create up to 4 copies of each archived data file. The system is currently configured to create the first (primary) copy of each archived file on the DISC 517 MO jukebox, and will be configured to create an additional copy of each file on the AIT tape jukebox. The AIT tapes will be stored off-line as they are filled.

All new data arriving at the NCEDC are being archived on the new mass storage system. Selected portions of the older datasets are being migrated to the new system in order to

return off-line data to on-line status. We plan to migrate the rest of the older data that is archived on the two Sony 300 GByte WORM jukeboxes to the new mass store after it has been upgraded to 2.6 TByte capacity this next year.

For increased availability and data reliability, most of the filesystems on the Ultra 450 are configured as mirrored (RAID level 1) filesystems with a available hot spare disk. Although this effectively halves the available disk space on the system, it provides the best protection for the data against disk failure.

The new mass storage system is currently housed in our temporary facilities in the Wellman trailers, since no space was available in McCone Hall. The system is on-line and archiving data, and will be moved to our new facilities in McCone Hall in early 1999. Table 1 shows the size of the various datasets at the NCEDC.

Joint Northern California Earthquake Catalog

Currently both the USGS and BSL construct and maintain earthquake catalogs for northern and central California. The "official" UC Berkeley earthquake catalog begins in 1910, and the USGS "official" catalog begins in 1966. Both of these catalogs are archived and available through the NCEDC, but the existence of 2 catalogs has caused confusion among both researchers and the public. The BSL and the USGS have spent considerable effort over the past year to define procedures for merging the data from the two catalogs into a single northern and central California earthquake catalog, in order to present a unified view of northern California seismicity. The differences in time period, variations in data availability, and mismatches in regions of coverage all complicate the task.

From 1910 through 1967, the BSL catalog is the primary source of northern California earthquake information. Only limited phase data are available for this time period, although location and magnitude information are provided.

The NCSN began to come on-line in 1966, and observations from this network are available beginning in July of that year. From this time forward, the BSL and USGS are working to generate a "joint" catalog by merging phase data and relocating the earthquakes. One of the initial complications in this project is matching up events between the two catalogs. Due to the sparse nature of the BSL instrumentation over the years, the BSL catalog is only complete at the magnitude 3 level while the USGS catalog is generally complete at the magnitude 2 level. However, the BSL catalog includes regional events from southern California, Nevada, Utah, Oregon, and Washington.

Thus neither catalog is a subset of the other. Other complications include foreshock/aftershock sequences, where one organization might read a foreshock and the mainshock and the other might read the mainshock and an immediate aftershock. Since limited phase data are available for the BDSN until 1976, most events during this period will combine the USGS location with the BSL magnitude. Where BSL phase data are available, an event that appears in both catalogs will have its phase and amplitude data merged, the event will be relocated with the combined phase data, and magnitudes will be recomputed using the available amplitude readings and new location.

The process of consolidating the BSL data to be merged into the joint catalog has uncovered many details in the original catalog which were ambiguous or poorly documented or inconsistent over time, such as the use of channel names for phase and amplitude readings. Significant time has been spent resolving these issues before the joint catalog is actually constructed.

Database Development

Most of the parametric data archived at the NCEDC, such as earthquake catalogs, phase and amplitude readings, waveform inventory, and instrument responses have been stored in flat text files. Flat file are easily stored and viewed, but are not efficiently searched. Over the last year, the BSL, in collaboration with the USGS and Caltech, has developed database schemas to store the parametric data from the joint earthquake catalog, station history, complete instrument response for all data channels, and waveform inventory.

The parametric schema will be used to define the tables and associations for the joint earthquake catalog. It allows for multiple hypocenters per event, multiple magnitudes per hypocenter, and association of phases and amplitudes with multiple versions of hypocenters and magnitudes respectively. The instrument response schema will represent full multi-stage instrument responses (including filter coefficients) for the broadband dataloggers. The hardware tracking schema will represent the interconnection of instruments, amplifiers, filters, and dataloggers over time. We plan to initially populate the database with the joint northern California earthquake catalog, BSL station history and instrument responses, and the BSL waveform inventory. The second stage of data will include the NCSN waveform inventory and later the NCSN instrument response data as it is made available. Additional details on the joint catalog effort and database schema development may be found at:

<http://quake.geo.berkeley.edu/>

Data Availability

The NCEDC continues to use the World Wide Web as a principal interface for users to request, search, and receive data from the NCEDC. The NCEDC has implemented a number of useful and original mechanism of data search and retrieval using the World Wide Web, and are available to anyone on the Internet. All of the documentation about the NCEDC, including the research users' guide, is available via the Web.

Users can perform catalog searches and retrieve hypocentral information and phase readings from the various earthquake catalogs at the NCEDC via easy-to-use forms on the Web. In addition, users can peruse the index of available broadband data at the NCEDC, and can request and retrieve broadband data in standard SEED format via the Web.

Access to all datasets is available via research accounts at the NCEDC. Information on requesting research accounts is available on the NCEDC web pages, <http://quake.geo.berkeley.edu/>.

For additional information on the NCEDC or its data sets, contact:

Douglas Neuhauser
510-652-0931
doug@seismo.berkeley.edu
or
dcmgr@quake.geo.berkeley.edu